Current Biology 24, R288–R294, March 31, 2014  http://dx.doi.org/10.1016/j.cub.2014.01.056

# Supergenes and Complex Phenotypes

Review

Tanja Schwander[1,*], Romain Libbrecht[2], and Laurent Keller[1,*]

**Understanding the molecular underpinnings of evolutionary adaptations is a central focus of modern evolutionary biology. Recent studies have uncovered a panoply of complex phenotypes, including locally adapted ecotypes and cryptic morphs, divergent social behaviours in birds and insects, as well as alternative metabolic pathways in plants and fungi, that are regulated by clusters of tightly linked loci. These 'supergenes' segregate as stable polymorphisms within or between natural populations and influence ecologically relevant traits. Some supergenes may span entire chromosomes, because selection for reduced recombination between a supergene and a nearby locus providing additional benefits can lead to locus expansions with dynamics similar to those known for sex chromosomes. In addition to allowing for the co-segregation of adaptive variation within species, supergenes may facilitate the spread of complex phenotypes across species boundaries. Application of new genomic methods is likely to lead to the discovery of many additional supergenes in a broad range of organisms and reveal similar genetic architectures for convergently evolved phenotypes.**

## Introduction

Supergenes are tight clusters of two or more loci each affecting a different developmental or behavioural characteristic. In combination, they provide integrated control of complex adaptive phenotypes segregating within species. The best known examples of supergenes are the morphologically and genetically distinct sex chromosomes, which have evolved independently in many groups of animals and plants [1,2]. Sex chromosomes typically evolve from autosomes via a mutation that causes its bearers to preferentially develop into a male or a female. Restricted recombination between the proto-X and proto-Y chromosomes then evolves rapidly so as to prevent recombination between genes with primary sex-determining roles, which would produce less fit organisms such as neuters or hermaphrodites [2,3].

The existence of supergenes outside of sex chromosomes was originally predicted by Ronald A. Fisher for Batesian mimicry systems in butterflies [4], where a harmless species experiences a low predation risk because it resembles a harmful model species avoided by predators. A single harmless species may imitate multiple, phenotypically divergent model species hence requiring specific trait combinations for matching each model. Fisher predicted that there should be strong selection for mechanisms suppressing recombination between loci affecting different traits when mixed trait combinations are detrimental. Similar reasoning predicted supergenes (referred to as 'co-adapted gene complexes') facilitating adaptation to local habitat conditions in species with broad distribution ranges [5,6] and motivated the development of theory investigating the conditions favouring the evolution of linkage [7–10]. Overall, these studies showed that the evolution of supergenes is expected to be adaptive when it causes linkage of beneficial epistatic interactions between alleles [6] or sets of locally adapted alleles [11,12].

The first empirical evidence that supergenes can function as a simple switch between trait combinations was provided for the 'pin' and 'thrum' floral types in *Primula vulgaris* and *Fagopyrum esculentum* (floral heteromorphy; Table 1) [13,14]. The differentiation of the two floral types is regulated by two loci in the supergene, one affecting length of the style, the female part of the flower, the other the position of the male anthers. Pin flowers have a long style reaching the opening of the floral tube and anthers located at the base of the floral tube, while thrum flowers have a short style and anthers at the opening of the floral tube (Figure 1). Pollinators reaching for nectaries at the floral tube base carry pollen from basal anthers on their head and pollen from anthers at the floral tube opening on their abdomen. Hence pollen from basal anthers preferentially lands on short styles while pollen from anthers at the floral tube opening reaches long styles. The position of anthers and styles thus reduces self-pollination and facilitates the reciprocal pollination between flower types. Self-fertilization is further prevented by a self-incompatibility system comprising two additional loci also located in the supergene.

Early studies also focussed on chromosomal inversions involved in local adaptation and speciation [15,16], because inversions can generate supergenes and are sometimes visible in karyotype surveys. However, direct proof for inversions maintaining supergene architecture was obtained only recently thanks to the advent of genome-wide sequencing techniques: two nested inversions were found to cause linkage among loci in the long-suspected butterfly mimicry supergene [17–19]. Inversions are, however, not the only mechanism that can maintain linkage between loci in a supergene (Table 1). Recent studies have led to the discovery of new linkage mechanisms and also a panoply of new phenotypes that are regulated by supergenes (Figure 1, Table 1). In addition to generating locally adapted ecotypes and cryptic morphs, tightly linked gene clusters are now known to be involved in divergent social behaviours in birds [20] and insects [21], and underlie the biosynthesis of chemical defence compounds in plants [22,23] and alternative metabolic pathways in fungi [24]. In this review, we discuss the mechanisms allowing, and the conditions favouring, the emergence of supergenes. We also argue that supergenes are more widespread than commonly appreciated and suggest that they can have important and unacknowledged consequences in convergent evolution across species.

## The Emergence of a Supergene

The advantage of supergene architectures is evident when loci affecting separate complementary phenotypic traits are tightly linked, thus preventing allele combinations that lead to non-optimal intermediate phenotypes. However, understanding the evolutionary mechanisms that generate these co-adapted gene sets, as well as how the production of unfit recombinant forms is avoided, remains a substantial challenge [10,25]. This question of how loci assemble into supergenes is related to a classical problem in evolutionary

[1]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland. [2]Laboratory of Insect Social Evolution, Rockefeller University, New York 10065, USA.
*E-mail: tanja.schwander@unil.ch (T.S.), laurent.keller@unil.ch (L.K.)

Table 1. Representative set of supergenes outside the well-documented sex chromosome systems.

| Trait (locus name) | Multiple origins? | Spread via hybridization? | Frequency | Linkage mechanism | Polymorphism maintenance | References |
|---|---|---|---|---|---|---|
| Floral heteromorphy (S-Locus) | Yes | ? | > 5 plant species[1] | Physical proximity (and ev. repetitive elements) | Self-incompatibility prevents production of homozygotes | [13,14,54] |
| Self-incompatibility (S-Locus) | Yes | ? | In species of half the angiosperm families[2] | Taxon-specific[3]: -Centromere location -Physical proximity -Structural variation | Self-incompatibility prevents production of homozygotes | [40–42,46,55,56] |
| Autosomal drive (t-haplotype, SD, GE, SK) | Yes | Yes | Widespread in different taxa[4] | Inversion(s) | Recessive lethals in the driver haplotypes | [26,34–36] |
| Asexual reproduction via seeds "Apomixis" (ASGR) | Yes | Yes | > 27 plant species[5] | Extreme structural differences (hemizygosity) | Unknown | [37,57–60] |
| Plumage polymorphism (ZAL2) | No | No | One bird *Zonotrichia albicollis* | Inversions | Disassortative mating between alternative morphs | [20,61] |
| Shell color polymorphism (*Cepea*/*Partula* supergene) | Yes | ? | 5 snail species[6] | Unknown | Probably spatially heterogeneous selection | [47,62,63] |
| Social organization (SB/Sb) | Probably[7] | Probably | ?[7] | Inversion | Recessive lethals in one haplotype; possibly spatially heterogeneous selection | [21,64] |
| Plant ecotypes (*Mimulus:* AN/PE, *Helianthus*: islands of divergence) | Yes | ? | Many cases, best described ones: *Mimulus* monkeyflowers, *Helianthus* sunflowers | Taxon-specific: -Chromosomal rearrangements -Inversions | Spatially heterogeneous selection | [39,65–67] |
| Cryptic female morph (OB-W) | Yes[8] | Yes | > 20 cichlid species | Physical proximity?[9] | Sex specificity and lack of OB/OB individuals in natural populations (unknown reasons) | [29] |
| Batesian mimicry (P-locus, H-locus) | Yes | Yes | > 5 species[10] | Inversions | Spatially heterogeneous selection | [17–19,68–71] |

[1]At least in *Primula sinensis*, *P. vulgaris*, *P. obconica*, *Turnera subulata*, *Fagopyrum esculentum*; no genetic information on additional species with floral heteromorphy.
[2]Best described cases in Brassicaceae, Solanaceae, Papaveraceae, Rosaceae, and Scrophulariaceae; other groups with little or no information on molecular basis of self-incompatibility.
[3]Location close to centromere in some Solanaceae and Scrophulariaceae; physical clustering and structural variation among haplotypes in Brassicaceae and some Rosaceae.
[4]At least seven well described cases including *Drosophila melanogaster* SD; *Mus musculus* t-haplotype (ssp *musculus, domesticus, castaneus, bactranus*) and cases in fungi and plants.
[5]Best described cases: *Pennisetum* (16 sp), *Cenchrus* (2 sp), *Panicum* (1sp).
[6]One case in four related species (*Cepea nemoralis*, *C. hortensis*, *C. vindobonensis*, *C. sylvatica*) and a different one in *Partula taeniata*.
[7]Only one ant (*Solenopsis invicta*) species with molecular characterization, many species with convergent phenotypes but unknown genetics.
[8]At least three independent origins in the lake Malawi Cichlid genera *Labeotropheus* and *Metriaclima*.
[9]Not due to inversions or structural rearrangements.
[10]*Heliconius numata*, *Papilio dardanus*, *Papilio polytes*, *Papilio memnon*, *Hypolimnas mysippus*; seven sympatric morphs that mimic *Melinea* butterflies in *H. numata*; mimetic morphs in other species (sometimes female-limited).

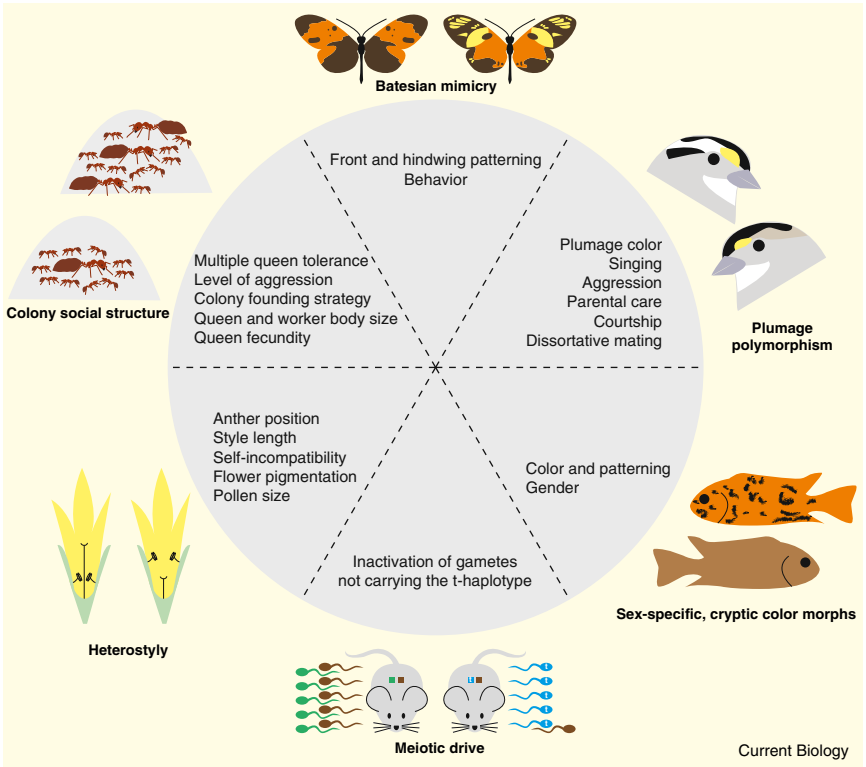


Figure 1. Supergenes regulate a wide spectrum of complex alternative phenotypes.

Alternative phenotypes are often differentiated by a suite of morphological, physiological and behavioral traits as nicely illustrated by the different social organizations in fire ants, the plumage polymorphism in white-throated sparrows, the Batesian mimicry morphs in *Heliconius* butterflies, and the female-specific, orange-blotch morphs in some cichlids (the most relevant traits affected by the supergene are listed in the grey circle for each example). More subtle phenotypes include the t-haplotype meiotic drive system in mice, intensively studied at the molecular level, and heterostyly in flowers that prevents self-fertilization and inbreeding.

biology: how can new beneficial traits requiring more than one novel mutation emerge in natural populations?

There are several well-known supergenes where linkage between multiple loci has been shown to be essential. For example, many meiotic drive systems require linkage between a 'killer' locus and a 'resistant responder'. One such system is the t-haplotype in mice (Table 1) [26]. In males heterozygous for t, ~90% of the eggs are fertilised by sperm carrying the t-haplotype instead of the expected 50% in the absence of drive. Increased transmission occurs because the killer locus in the t-haplotype damages wild-type sperm. Sperm carrying the t-haplotype are not damaged because the killer locus is linked to an allele at the *Tcr* (t-complex responder) locus, which provides protection from the damaging effects of the driver (via an unknown mechanism). The *Tcr* locus is expressed in a haploid-specific manner such that only sperm carrying it are rescued. Importantly, the killer locus can only become established in a population if linked to the drive-resistant *Tcr* allele. If recombination unlinks the resistant allele from the killer locus, sperm carrying the killer locus die, as they are no longer immune to the killer locus' effect. Interestingly, the killer effect is due to the additive action of three or more loci, but the ancestral t-haplotype most likely comprised only one or a few of the current killing loci. Once the original t-supergene was formed by linking *Tcr* with a killing locus, additional loci increasing drive were recruited to the linkage-block, resulting in a progressive expansion of the supergene, which currently spans over ~40 Mb on mouse chromosome 17.

Drive systems such as the t-haplotype require at least partial linkage among the core components for their initial evolution [27,28]. The degree of linkage among loci depends on the physical distance between them and on the recombination rate in the chromosome region where they are located. There are three main mechanisms that can lead to clustering of loci involved in adaptive supergene structures (Figure 2A).

The first mechanism is a new mutation leading to beneficial interactions with another closely located locus. While this is probably the simplest mechanism, there is currently no confirmed example, most likely because information pertaining to the order in which mutations at different loci appear is difficult to obtain from natural populations. It is therefore currently impossible to determine whether or not this mechanism is a common route to the origin of supergenes. One possible example is the orange-blotch (OB) phenotype in some cichlid species of Lake Malawi (Table 1) [29]. The OB phenotype is advantageous in females because it provides an alternative form of crypsis by matching the background of the environment that the cichlids inhabit. By contrast, it is disadvantageous in males, because it disrupts their nuptial colour and thus decreases their probability to find a mate. Crosses designed for mapping the OB locus revealed that the locus also functions as a novel female sex determiner in at least one species. Thus, a female-determining mutation would have occurred close to the OB locus. This scenario, which still awaits confirmation through molecular studies, thus seems to have generated an adaptive supergene haplotype by causing female-limited transmission of the OB phenotype.

The second mechanism is a duplication generating a novel gene that interacts with another closely located gene. Repeated gene duplications have generated the major histocompatibility complex (MHC), which in humans spans >3 Mb on the short arm of chromosome 6 and is characterized by reduced recombination. About 40% of the genes in this region participate in immune responses [30]. Although MHC is not generally considered a supergene, it is likely that beneficial interactions among specific alleles at different loci have selected for increased linkage to maintain particular haplotypes [31].

The third possible mechanism generating clusters of interacting genes is by translocation. Comparative genomic analyses have revealed several cases of gene clusters implicated in the biosynthesis of defence compounds in plants that evolved via intra-genomic gene translocations [32,33].

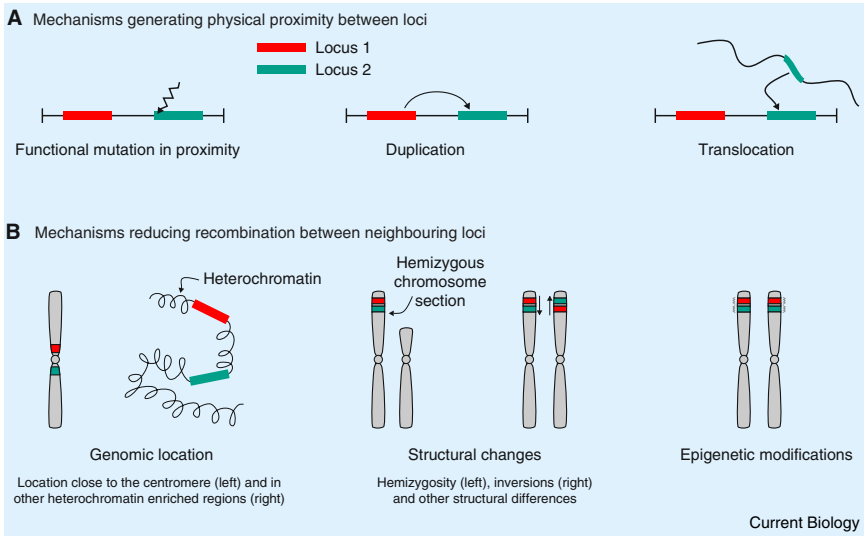


Figure 2. Emergence of a supergene requires physical linkage between loci.

Different mechanisms can generate proximity between loci (A) and suppress recombination between them (B). New mutations, duplications and translocations can generate coadapted gene complexes. The degree of linkage among loci in these complexes depends on their location in the genome and can be enhanced via structural variation (including inversions) and epigenetic modification.

This contrasts with the majority of biosynthetic pathways (including the anthocyanin synthesis pathway in *Zea mays* and other well characterized secondary metabolic pathways), which comprise genes that are randomly distributed in the genome. The clustering of loci encoding components of the same pathway ensures that all components are transmitted together. Importantly, a failure to transmit the complete cluster would lead not only to reduced resistance to diseases and pests but also to the accumulation of toxic intermediates. For example, the accumulation of triterpene defence pathway intermediates has been shown to impede normal growth and development in oat and *Arabidopsis thaliana* [32].

## Suppression of Recombination in the Supergene

A supergene is defined by the fact that it is inherited as one unit, rather than as individual components. This means that recombination between the loci in a supergene needs to be minimised. While close physical location of loci is important, several additional mechanisms can further decrease the probability of recombination (Figure 2B); the three most important ones being outlined below.

The first factor affecting the recombination rate between physically close loci is the location in the genome. Some parts of the genome, such as centromeres and other hetero-chromatic regions, as well as regions including high density of interspersed repeats, are characterized by reduced recombination rates. Several known supergenes are located near the centromere (Table 1), suggesting that genomic regions with low recombination are favourable for their emergence. For example, meiotic drive systems acting by killing gamete copies not containing them are frequently located near the centromere. Such examples include the mouse t-haplotype [26], the *Drosophila melanogaster* Segregation Distorter (SD) system [34], as well as the less-studied gamete-killing systems in the plants *Lycopersicon esculentum* and *L. pimpinellifolium* (Gamete Eliminator locus) [35] and the Spore-Killing (SK) alleles in several fungi [36] (Table 1).

A second important factor that may decrease recombination between loci are structural differences between homologous chromosomal regions. In the most extreme cases, supergenes are located in large genomic fragments that are entirely absent in the wild type. This occurs, for example, in plants characterized by asexual reproduction through seeds (Table 1). Asexual seed production is caused by the combined action of a set of loci in the distal region of a chromosome arm referred to as the apospory-specific genomic region (ASGR) [37]. At least one locus controls the production of unreduced gametes, a different one the fertilization-independent development of these gametes, and in some cases additional loci regulate fertilization-independent endosperm development. Plants producing asexual seeds still produce meiotically reduced pollen. During the production of pollen, the loci in the ASGR remain as one set because the chromosome homologous to the one carrying the ASGR does not possess this distal region, hence ASGR is always present in only one copy (hemizygous), precluding recombination. ASGR-carrying pollen can fertilize sexual eggs and generate new plant clones. As a consequence, ASGRs segregate in sexual plant populations and have spread repeatedly across species barriers. Interestingly, ASGRs have evolved convergently in several plant families via different molecular pathways, but the processes underlying this structural convergence remain enigmatic [37].

Structural differences can also be caused through chromosomal inversions. Inversions are fundamental features of some of the best-studied supergene systems. In the classical case of butterfly mimetic morphs in *Heliconius* butterflies, two non-overlapping chromosome inversions generated three chromosomal arrangements for the P supergene, each corresponding to a specific wing-pattern phenotype [17,18]. Similarly, an inversion caused two non-recombining chromosomal arrangements (AN or PE) in the yellow monkeyflower *Mimulus guttatus* that underlie alternative development of inland annual and coastal perennial ecotypes, which also differ in flowering time and other morphological traits [38,39].

In some cases, inversions can be associated with recombination suppression over significant fractions of a chromosome. For example, in the fire ant *Solenopsis invicta* a supergene regulating the presence of one or more queens per nest, and a suite of associated traits in queens and workers, spans 13 Mb (55%) of a chromosome [21]. Recombination between the two variants (SB and Sb) of the supergene is completely absent, most likely because of one large (>9 Mb) and maybe additional smaller inversions. Similarly, a plumage polymorphism in the white-throated sparrow (*Zonotrichia albicollis*) is associated with alternative chromosomal arrangements (ZAL2 and ZAL2$^{m}$) caused by two pericentric inversions, spanning at least 98 Mb (86%) of

the second chromosome [20]. The alternative morphs determined by ZAL2 haplotypes are also characterized by different behavioural syndromes, but it remains to be confirmed that the behavioural traits are regulated by loci within the supergene. An important behavioural trait is mate preference; indeed individuals of each morph have a strong preference for mates of the alternative morph, which is likely fundamental to the maintenance of the supergene polymorphism in this species (see below).

Less pronounced structural differences between chromosomes have also been proposed to decrease recombination rates. For example, it is believed that recombination between haplotypes at a supergene causing self-incompatibility in Brassicaceae plants (S-locus) is suppressed because different haplotypes are characterized by different lengths and gene arrangements, and vary in gene content [40]. The S-locus prevents self-pollination and favours outcrossing via its two core components, one expressed on the pollen coat and one expressed on the stigma surface [41]. If the two components come from the same S-haplotype, fertilization is prevented via a cascade triggered by their interaction. Recombination between different S-haplotypes would allow for self-pollination in individuals with the recombined haplotype and therefore cause the breakdown of the self-incompatibility system [42]. Self-incompatibility systems in plants have evolved several times independently, via different molecular mechanisms and with different factors suppressing recombination between haplotypes (Table 1).

When reduced recombination is associated with structural differences between haplotypes, it is often difficult to disentangle whether these differences initially prevented recombination, or whether they accumulated secondarily. Indeed the cessation of recombination between homologous chromosome regions should facilitate the structural divergence between these regions, via synteny change and gene copy number divergence. Thus, structural differences between supergene haplotypes may emerge as a consequence of reduced recombination rather than causing it.

The final and least understood way to suppress recombination is via epigenetic modifications (Figure 2B). Epigenetic changes, especially DNA methylation, can influence the frequency and distribution of recombination events along chromosomes, often via modifications of chromatin and histone organisation [43,44]. Epigenetic mechanisms thus have the potential to specifically suppress recombination between individual loci in a genomic region, in contrast to most structural changes that will tend to suppress recombination between many loci. Epigenetic modifications are believed to be involved in recombination suppression in the sex-determining region of certain sex chromosomes [45]. Similarly, epigenetic changes may act in addition to structural differences to suppress recombination between S-haplotypes in Brassicaceae [46].

## Supergenes Doomed for Extinction?

The evolutionary fate of supergenes depends on several factors. In the short term, the dynamics of supergene haplotypes depend on the fitness of the phenotypes they regulate. Genetically-determined alternative phenotypes segregating within species should have equal fitness if they are to coexist — a classical problem in ecological genetics [6,15]. New supergenes are likely to either have a higher or a lower fitness than the wild-type haplotype, leading to their fixation or elimination. However, a stable polymorphism can be maintained under certain selection regimes. This would be, for example, the case under spatial heterogeneity whereby alternative genetic variants are each favoured in different environments. Such a situation has been suggested to maintain supergene polymorphism in *Mimulus* monkeyflowers, where the perennial ecotype is widespread in costal habitats and the annual ecotype occurs in more ephemeral inland habitats [14]. Locally co-adapted alleles within the supergene affect leaf and flower morphology, flowering time as well as reproductive isolation between the perennial and annual ecotypes. Spatial heterogeneity is also the most likely explanation for the maintenance of a polymorphism at the supergene controlling shell patterning in snails (Table 1). In the example, different loci in the supergene regulate shell background colour, banding presence, and the number and width of bands, whereby the right combination of traits generates a locally cryptic shell pattern [47].

The main mechanism maintaining supergene polymorphisms, however, seems to be negatively frequency-dependent selection (Table 1). The strongest frequency dependence probably occurs for sex chromosomes, given reproduction is only possible between individuals of opposite sex. Another good example of frequency-dependent selection is that of self-incompatibility systems in plants. If one S-haplotype reaches high frequencies, individuals carrying this haplotype may lack compatible mating partners whereas rare haplotypes in the same population would be compatible with the majority of individuals. Such local mate limitation would prevent fixation of the frequent haplotype and favour rare ones [48]. Frequency-dependent selection also occurs in the case of the white-throated sparrows, where individuals of each morph display strong preference for mates of the alternative morph [20], again generating higher relative success for the rare morph.

Finally, it is important to note that, in many cases, an important factor leading to polymorphism maintenance is the presence of recessive lethal alleles in one of the supergene haplotypes (Table 1). Recessive lethals can maintain polymorphism when they are linked to the selectively favoured supergene variant. A stable polymorphism will be achieved when positive selection for this variant is compensated by the mortality of homozygous individuals. The presence of recessive lethals is generally considered to be a consequence of the lack of recombination for haplotypes that never occur in a homozygous state. The accumulation of deleterious mutations, including recessive lethals, is well recorded in the context of the evolution of the Y (and W) sex chromosomes, which tend to degenerate in many species (recently reviewed in [49]). Y-chromosomes are never in a homozygous state because they only occur in XY males (the same is true for W-chromosome in WZ females). Some supergenes, notably those associated with self-incompatibility, are also expected to occur exclusively as heterozygotes in the population. By contrast, for supergenes not associated with mating types, occasional recombination between copies of the same haplotype in homozygotes should allow for a reduction of the genetic load. Thus, the presence of lethals in early stages of the evolution of supergenes may have acted as potent mechanisms maintaining stable polymorphisms by preventing their fixation via lethality of the homozygotes. Importantly, once a supergene haplotype contains one recessive lethal, there will be no more selection against the accumulation of additional recessive mutations.

This may explain why there seems to be a dynamic of accumulation of lethal recessive alleles, as in the well-studied case of the t-haplotype.

## Conclusions

The highlighted case studies illustrate a wide diversity of adaptive phenotypes regulated via supergenes. In spite of this phenotypic diversity, there appear to be two common patterns shared among different supergene systems. First, for most traits known to be regulated by supergenes, the same traits have evolved multiple times independently in different species (Table 1), frequently via co-option of different molecular mechanisms. Second, traits regulated via supergenes appear to often spread across species boundaries via hybridization (Table 1), introducing complex adaptations into new genetic backgrounds.

In addition to spreading supergenes across genetic backgrounds, we suggest that hybridization between differentiated populations or species could directly generate novel supergenes. Molecular divergence between genomic regions greatly reduces the local rate of recombination [50], due to control mechanisms restricting recombination to homologous regions. Gene flow could therefore lead to the introgression of genome segments that freely recombine in the original genetic background, but experience suppressed recombination in a new genetic background. While the current data do not allow one to test whether known cases of supergenes evolved via historical hybridization, we expect this mechanism to be at the origin of many polymorphic supergene complexes, given its simplicity and the ubiquity of sporadic gene flow between species.

The abundance and diversity of complex phenotypes regulated by supergenes highlights the role that genomic structural variation can have in adaptation and speciation. Genome-wide sequencing has already revealed several unexpected cases of alternative phenotypes being regulated by supergene haplotypes. Many additional cases are likely to be uncovered in the near future, especially for complex traits segregating as expected for a single Mendelian locus. Several of these are already known to be associated with polymorphic inversions and adaptive variation [51]. The inversion polymorphism at the 17q21.31 locus in humans, for which women carrying a specific haplotype have more children than non-carriers in a Northern European population [52], nicely illustrates such a case. Supergenes would be those cases where further molecular studies will uncover that adaptive variation is caused by more than one locus in the inversion.

There are important challenges in the study of potential (and confirmed) supergenes. The first is to identify the genes underlying the production of alternative phenotypes. Since there are typically many loci in supergenes, and because they remain linked together, it has proven very difficult to determine which of them are causal for the alternative phenotypes. Another important task will be to distinguish elements present during the initial spread of the phenotype from those recruited at later stages. A potential avenue for such studies will be to consider cases where the same supergenes occur in several closely related species. Comparative genomic analyses of these species in a phylogenetic context should allow reconstructing the history of a supergene, in a similar way as it has been done for Y-chromosomes in mammals [53]. More generally, supergenes provide contrasting and complementary systems to study the evolution of non-recombining genome regions and the evolution of highly divergent phenotypes, be they alternative morphs or sexes. It will be of great interest to determine how frequently dimorphic phenotypes in a given species are produced purely by phenotypic plasticity, by polymorphic regulatory elements affecting the expression of several genes, or by supergenes whose sequence differences are directly responsible for the alternative phenotypes. Our prediction is that supergenes are implicated in many more cases than currently recognized.

### Acknowledgements

We thank Nicolas Perrin, Jessica Purcell and two anonymous reviewers for their useful comments. T.S. was supported by a grant from the Swiss NSF, R.L. by a Marie Curie postdoctoral fellowship, and L.K. by several grants from the Swiss NSF and an advanced ERC grant.

### References

1. Bull, J.J. (1983). Evolution of Sex Determining Mechanisms (Menlo Park, CA: Benjamin/Cummings).
2. Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. Curr. Biol. *6*, 149–162.
3. Lewis, D. (1942). The evolution of sex in flowering plants. Biol. Rev. *17*, 46–67.
4. Fisher, R.A. (1930). The Genetical Theory of Natural Selection (Oxford: Clarendon Press).
5. Dobzhansky, T. (1948). Genetics of natural populations. XVII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographical regions. Genetics *33*, 588–602.
6. Dobzhansky, T. (1970). Genetics of the Evolutionary Process (New York: Columbia University Press).
7. Clarke, C.A., and Sheppard, P.M. (1960). The evolution of mimicry in the butterfly *Papilio dardanus*. Heredity *14*, 163–173.
8. Clarke, C.A., and Sheppard, P.M. (1960). Super-genes and mimicry. Heredity *14*, 175–185.
9. Turner, J.R.G. (1967). On supergenes I. Evolution of supergenes. Am. Nat. *101*, 195–221.
10. Charlesworth, B., and Charlesworth, D. (1975). Theoretical genetics of batesian mimicry II. Evolution of supergenes. J. Theor. Biol. *55*, 305–324.
11. Kirkpatrick, M., and Barton, N. (2006). Chromosome inversions, local adaptation and speciation. Genetics *173*, 419–434.
12. Kirkpatrick, M. (2010). How and why chromosome inversions evolve. PLoS Biol. *8*, e1000501.
13. Mather, K. (1950). The genetical architecture of heterostyly in *Primula sinensis*. Evolution *4*, 340–352.
14. Gilmartin, P.M., and Li, J. (2010). Homing in on heterostyly. Heredity *105*, 161–162.
15. Ford, E.B. (1971). Ecological Genetics, 3rd Edition (London: Chapman & Hall).
16. White, M.J.D. (1978). Modes of Speciation (San Francisco: W. H. Freeman).
17. Joron, M., Papa, R., Beltran, M., Chamberlain, N., Mavarez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S.J., Rogers, J., *et al*. (2006). A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. PLoS Biol. *4*, e303.
18. Joron, M., Frezal, L., Jones, R.T., Chamberlain, N.L., Lee, S.F., Haag, C.R., Whibley, A., Becuwe, M., Baxter, S.W., Ferguson, L., *et al*. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature *477*, 203–206.
19. Jones, R.T., Salazar, P.A., ffrench-Constant, R.H., Jiggins, C.D., and Joron, M. (2012). Evolution of a mimicry supergene from a multilocus architecture. Proc. Roy. Soc. B. *279*, 316–325.
20. Thomas, J.W., Caceres, M., Lowman, J.J., Morehouse, C.B., Short, M.E., Baldwin, E.L., Maney, D.L., and Martin, C.L. (2008). The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. Genetics *179*, 1455–1468.
21. Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.C., Shoemaker, D., and Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. Nature *493*, 664–668.
22. DellaPenna, D., and O'Connor, S.E. (2012). Plant science. Plant gene clusters and opiates. Science *336*, 1648–1649.
23. Takos, A.M., and Rook, F. (2012). Why biosynthetic genes for chemical defense compounds cluster. Trends Plant Sci. *17*, 383–388.
24. Brakhage, A.A. (2013). Regulation of fungal secondary metabolism. Nat. Rev. Microbiol. *11*, 21–32.
25. Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc. Natl. Acad. Sci. USA *110*, E1743–E1751.

26. Lyon, M.F. (2003). Transmission ratio distortion in mice. Annu. Rev. Genet. *37*, 393–408.
27. Charlesworth, B., and Hartl, D.L. (1978). Population dynamics of the segregation distorter polymorphism of *Drosophila melanogaster*. Genetics *89*, 171–192.
28. Crow, J.F. (1991). Why is Mendelian segregation so exact. BioEssays *13*, 305–312.
29. Roberts, R.B., Ser, J.R., and Kocher, T.D. (2009). Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science *326*, 998–1001.
30. The MHC sequencing consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. Nature *401*, 921–923.
31. Beck, S., and Trowsdale, J. (2000). The human major histocompatibility complex: lessons from the DNA sequence. Annu. Rev. Genomics Hum. Genet. *1*, 117–137.
32. Chu, H.Y., Wegel, E., and Osbourn, A. (2011). From hormones to secondary metabolism: the emergence of metabolic gene clusters in plants. Plant J. *66*, 66–79.
33. Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T.R., Li, Y., Meade, F., Teodor, R., Vaistij, F.E., *et al*. (2012). A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science *336*, 1704–1708.
34. Larracuente, A.M., and Presgraves, D.C. (2012). The selfish Segregation Distorter gene complex of *Drosophila melanogaster*. Genetics *192*, 33–53.
35. Rick, C.M. (1971). Tomato Ge Locus - Linkage relations and geographic distribution of alleles. Genetics *67*, 75–85.
36. van der Gaag, M., Debets, A.J.M., Oosterhof, J., Slakhorst, M., Thijssen, J., and Hoekstra, R.F. (2000). Spore-killing meiotic drive factors in a natural population of the fungus Podospora anserina. Genetics *156*, 593–605.
37. Ozias-Akins, P., and van Dijk, P.J. (2007). Mendelian genetics of apomixis in plants. Annu. Rev. Genet. *41*, 509–537.
38. Hall, M.C., Basten, C.J., and Willis, J.H. (2006). Pleiotropic quantitative trait loci contribute to population divergence in traits associated with life-history variation in *Mimulus guttatus*. Genetics *172*, 1829–1844.
39. Lowry, D.B., and Willis, J.H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. PLoS Biol. *8*, e1000500.
40. Casselman, A.L., Vrebalov, J., Conner, J.A., Singhal, A., Giovannoni, J., Nasrallah, M.E., and Nasrallah, J.B. (2000). Determining the physical limits of the Brassica S locus by recombinational analysis. Plant Cell *12*, 23–33.
41. Iwano, M., and Takayama, S. (2012). Self/non-self discrimination in angiosperm self-incompatibility. Curr. Opin. Plant Biol. *15*, 78–83.
42. Sijacic, P., Wang, X., Skirpan, A.L., Wang, Y., Dowd, P.E., McCubbin, A.G., Huang, S., and Kao, T.H. (2004). Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature *429*, 302–305.
43. Maloisel, L., and Rossignol, J.L. (1998). Suppression of crossing-over by DNA methylation in *Ascobolus*. Genes Dev. *12*, 1381–1389.
44. Yelina, N.E., Choi, K., Chelysheva, L., Macaulay, M., de Snoo, B., Wijnker, E., Miller, N., Drouaud, J., Grelon, M., Copenhaver, G.P., *et al*. (2012). Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. PLoS Genet. *8*, e1002844.
45. Kondo, M., Nanda, I., Schmid, M., and Schartl, M. (2009). Sex determination and sex chromosome evolution: Insights from Medaka. Sexual Dev. *3*, 88–98.
46. Yang, Q., Zhang, D., Li, Q., Cheng, Z., and Xue, Y. (2007). Heterochromatic and genetic features are consistent with recombination suppression of the self-incompatibility locus in *Antirrhinum*. Plant J. *51*, 140–151.
47. Murray, J., and Clarke, B. (1976). Supergenes in polymorphic land snails I. *Partula taeniata*. Heredity *37*, 253–369.
48. Wright, S. (1939). The distribution of self-sterility alleles in populations. Genetics *24*, 538–552.
49. Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. *14*, 113–124.
50. Waldman, A.S. (2008). Ensuring the fidelity of recombination in mammalian chromosomes. BioEssays *30*, 1163–1171.
51. Hoffmann, A.A., Sgro, C.M., and Weeks, A.R. (2004). Chromosomal inversion polymorphisms and adaptation. Trends Ecol. Evol. *19*, 482–488.
52. Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G., *et al*. (2005). A common inversion under selection in Europeans. Nat. Genet. *37*, 129–137.
53. Li, G., Davis, B.W., Raudsepp, T., Pearks Wilkerson, A.J., Mason, V.C., Ferguson-Smith, M., O'Brien, P.C., Waters, P.D., and Murphy, W.J. (2013). Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res. *23*, 1486–1495.
54. Labonne, J.D., Tamari, F., and Shore, J.S. (2010). Characterization of X-ray-generated floral mutants carrying deletions at the S-locus of distylous *Turnera subulata*. Heredity *105*, 235–243.
55. De Nettancourt, D. (2000). Incompatibility and Incongruity in Wild and Cultivated Plants, 2 Edition (Berlin: Springer-Verlag).
56. Takayama, S., and Isogai, A. (2005). Self-incompatibility in plants. Annu. Rev. Plant Biol. *56*, 467–489.
57. Ozias-Akins, P., Akiyama, Y., and Hanna, W.W. (2003). Molecular characterization of the genomic region linked with apomixis in *Pennisetum/Cenchrus*. Funct. Integr. Genomics *3*, 94–104.
58. van Dijk, P.J. (2003). Ecological and evolutionary opportunities of apomixis: insights from *Taraxacum* and *Chondrilla*. Philos. Trans. R. Soc. Lond. B. *358*, 1113–1121.
59. van Dijk, P.J. (2009). Apomixis: Basics for Non-botanists (Dordrecht: Springer), pp. 47–67.
60. Akiyama, Y., Goel, S., Conner, J.A., Hanna, W.W., Yamada-Akiyama, H., and Ozias-Akins, P. (2011). Evolution of the apomixis transmitting chromosome in *Pennisetum*. BMC Evol. Biol. *11*, 289.
61. Thorneycroft, H.B. (1975). A cytogenetic study of the white-throated sparrow, *Zonotrichia albicollis* (Gmelin). Evolution *29*, 611–621.
62. Jones, J.S., Leith, B.H., and Rawlings, P. (1977). Polymorphism in *Cepaea* - Problem with too many solutions. Annu. Rev. Ecol. Syst. *8*, 109–143.
63. Cook, L.M. (1998). A two-stage model for *Cepaea* polymorphism. Phil. Trans. R. Soc. London B. *353*, 1577–1593.
64. Keller, L., and Parker, J.D. (2002). Behavioral genetics: a gene for supersociality. Curr. Biol. *12*, R180–R181.
65. Hall, M.C., Lowry, D.B., and Willis, J.H. (2010). Is local adaptation in Mimulus guttatus caused by trade-offs at individual loci? Mol. Ecol. *19*, 2739–2753.
66. Lowry, D.B., Hall, M.C., Salt, D.E., and Willis, J.H. (2009). Genetic and physiological basis of adaptive salt tolerance divergence between coastal and inland *Mimulus guttatus*. New Phytol. *183*, 776–788.
67. Renaut, S., Grassa, C.J., Yeaman, S., Moyers, B.T., Lai, Z., Kane, N.C., Bowers, J.E., Burke, J.M., and Rieseberg, L.H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Comm. *4*, 1827.
68. Clarke, C.A., and Sheppard, P.M. (1963). Interaction between major genes and polygenes in determination of mimetic patterns of *Papilio dardanus*. Evolution *17*, 404–413.
69. Joron, M., Wynne, I.R., Lamas, G., and Mallet, J. (1999). Variable selection and the coexistence of multiple mimetic forms of the butterfly *Heliconius numata*. Evol. Ecol. *13*, 721–754.
70. Nijhout, H.F. (2003). Polymorphic mimicry in *Papilio dardanus*: Mosaic dominance, big effects, and origins. Evol. Dev. *5*, 579–592.
71. Clark, R., Brown, S., Heckel, D., Jiggins, C.D., Collins, S., and Vogler, A.P. (2007). The molecular genetics of the H locus: colour polymorphism determination in Papilio dardanus. J. Insect Sci. *7*, 8–9.